The Method of Purging Applied to Repeated Cross-Sectional Data

Grotenhuis, Manfred te; Eisinga, Rob; Scheepers, Peer

Published in:
Quality & Quantity

DOI:
10.1023/B:QUQU.0000013238.96787.53

Publication date:
2004

Quality & Quantity **38:** 1–16, 2004.

The Method of Purging Applied to Repeated Cross-Sectional Data

Practical applications using logistic and linear regression analysis

MANFRED TE GROTENHUIS*, ROB EISINGA and PEER SCHEEPERS
Department of Social Science Research Methods, University of Nijmegen, PO Box 9104, 65000 HE Nijmegen, The Netherlands

Abstract. In cross-sectional survey research, it is quite common to estimate the (standardized) effect of independent variable(s) on a dependent variable. However, if repeated cross-sectional data are available, much is to be gained if the consequences of these effects on longitudinal social change are considered.

To assess these consequences, we describe a type of simulation in which longitudinal shifts in the independent variable's distribution, and longitudinal variation in effect on the dependent variable are 'purged' from the data. Although the method of purging is known for many years, we add new practical features by relating the method to logistic and linear regression analysis. Because both logistic and linear regression analysis can be found in all major statistical packages, the method of purging is made available to a wider group of social scientists. With the use of repeated cross-sectional data, gathered in the Netherlands between 1970 and 1998, the new practical features of the purging method are shown, using the SPSS package.

Key words: purging, simulation, counter factual analysis, repeated cross-sectional survey, logistic and linear regression analysis

Introduction

There is a long tradition in social sciences of collecting cross-sectional survey data. As a result, a massive quantity of longitudinal data is available nowadays. These data are in many events more or less suitable to test causal models. Besides, these data are often tailor-made to test to what extent parameters in these models vary over time. However, to explain longitudinal social change, one has to go beyond the causal modeling of effects and their over-time variation. To answer the question on the causes of social change, we have to asses the *consequences* of shifting distribution(s) and of varying effect(s). One way to assess these consequences, is to simulate a counter factual situation in which both the distribution of the independent variable(s) and its effect(s) remain constant over time. This kind of simulation is often labelled 'purging' (for an overview see: Clogg, 1978; Clogg &

* E-mail: m.tegrotenhuis@maw.kun.nl

Table I. Relationship between church membership and confessional party voting in 1970 and 1998, frequencies and (percentages)

Vote for confessional party	Church membership: 1970 Non-member	1970 Church member	1970 Total	1998 Non-member	1998 Church member	1998 Total
No	472	286	758	860	351	1211
	(95.7)	(37.9)	(60.7)	(96.1)	(58.5)	(81.0)
Yes	21	469	490	35	249	284
	(4.3)	(62.1)	(39.3)	(3.9)	(41.5)	(19.0)
Total	493	755	1248	895	600	1495
	(39.5)	(60.5)	(100)	(59.9)	(40.1)	(100)

Data derived from the 'Cultural Changes' data-set collected by the Social and Cultural Planbureau (SCP), The Netherlands: author's calculations.

Eliasen, 1988; Clogg et al., 1990) because one eliminates or purges the impact of shifting distribution(s) and varying effect(s).[1]

For example, in the Netherlands, the mainstream confessional political party CDA suffered from a massive loss of votes in the past decades. Also, there is abundance evidence that religious affiliation has a strong effect on voting behavior and over time this effect declined (Nieuwbeerta, 1995). Finally, there was a profound distributional shift, i.e., church membership rates were declining during the past century in the Netherlands.(Eisinga & Felling, 1990). These two facts, i.e., declined church membership rates and a declined effect of religious affiliation, may explain the over-time loss of confessional votes (Eisinga et al., 1997). In the next section we will use this example to explain and demonstrate the practical applications of purging using logistic and linear regression analysis.

Before we do so, we first introduce a 2×2 table based on two points in time, just to outline the basic idea behind the purging method applied to longitudinal data. Next, we will show how purging can be done using logistic regression analysis and a few constraints. Moreover, we will introduce more independent variables and a decomposition of the overall impact. Finally, we will show how the method of purging can be applied to a continuous dependent variable using a linear regression model. These practical extensions, i.e., using (multiple) logistic and linear regression analysis to purge distributional shift(s) and varying effect(s), are the core contribution of this article.

1. The Method of Purging Exemplified by Means of a 2×2 Table

In this section we will give a example of purging on a 2×2 table with data from repeated cross-sectional surveys. To save space, we restricted our data to two points in time, i.e., 1970 and 1998. The data from the two samples are shown in Table I.

Table II. Relationship between church membership and confessional party voting in 1998, constant association and constant distribution, frequencies and (percentages)

Vote for confessional party	Church membership:					
	Constant association in 1998			Constant association in 1998		
	Non-member	Church member	Total	Non-member	Church member	Total
No	857 (95.7)	227 (37.9)	1084 (72.5)	568 (96.1)	329 (58.5)	1097 (73.4)
Yes	38 (4.3)	373 (62.1)	411 (275)	23 (3.9)	375 (41.5)	398 (26.6)
Total	895 (59.9)	600 (40.1)	1495 (100)	591 (39.5)	904 (60.5)	1495 (100)

According to the row totals in Table I, (mainstream) confessional parties suffered from heavy losses between 1970 and 1998.[2] In 1970 39.3% of the Dutch voted for a confessional party, whereas in 1998 the percentage was down to 19.0. This political turnover is partly due to changes in church members' voting behavior, as can be seen from the interior (shaded) cells in Table I. According to these cells, 62.1% of all church members voted for a confessional party in 1970 but it declined to 41.5% in 1998. Confessional party voting among non-members was quite rare: only 4.3% of them did so in 1970 and 3.9% in 1998. The change in voting behavior of especially church members is reflected in the odds ratios as a measure of association: 36.9 in 1970 and 17.43 in 1998. Next to this change in association, we can also ascertain a profound shift in church membership rates. In 1970, 60.5% of the Dutch population considered themselves to be church members, whereas in 1998 this was 40.1%.

So, two processes may be responsible for the loss of confessional votes between 1970 and 1998. First, a change in the association between church membership and confessional party voting as church members changed their voting behavior. Second, a distributional shift in the (independent) variable church membership as the percentage non-members increased between 1970 and 1998.

To assess the impact of each of these two processes, we performed two simulations, one in which the association between church membership and confessional party voting was held constant, and one in which the church membership rates were held constant. The outcomes are presented in Table II.

In Table II, the left 2×2 tabulation represents the simulation in which the association between church membership and religious voting is held constant, i.e., the odds ratio is held on the 1970 level. To archieve this, we used the 1998 distribution of church membership (895/600) and the conditional percentages from 1970 (cf. Table I). On the basis of this information, we recalculated the cell frequencies for 1998. For instance in cell 0.0 (non-member, no vote for confessional party) the new frequency is $895 \times 0.957 = 857$. On the basis of the new frequencies, the percentage that did vote for confessional party is recalculated: $(38 + 373)/1495 \times 100 = 27.5\%$). So, if a constant association is simulated, the percentage of votes for confessional parties in 1998 is 27.5 instead of the observed 19.0 (cf. Table I) and lies well outside the 95% confidence interval around 19.0.[3] Practically, this finding implies that if the association between church membership and religious voting not had changed over time, the percentage of votes for a confessional party would have amounted to 27.5, which is significantly different from the observed percentage.

In the right part of Table II, we simulated constant distribution. To keep the distribution of the church membership variable in 1998 on the 1970 level, we use the 1970 distribution (39.5/60.5), as new column total percentages in 1998. On the basis of these percentages we recalculated column total frequencies for 1998. For instance, the number of non-members is $0.395 \times 1495 = 591$. On the basis of these frequencies we calculated new frequencies for the shaded cells. For example, the number of non-members who did not vote for a confessional party is $591 \times 96.1 = 568$. The purged percentage of people who voted for confessional parties in 1998 is $23 + 375/1495 \times 100 = 26.6\%$. So, if we simulate a constant distribution of church membership, the percentage of votes for confessional parties is 26.6 instead of the observed 19.0 and again lies outside the 95% confidence interval (cf. note 3). Practically, this finding implies that if the church membership's distribution not had changed over time, the percentage for a confessional party would have amounted to 26.6 and is significant different from the observed percentage. As these two simulations have an impact of respectively 8.5 and 7.6 (purged % in 1998 – observed % in 1998), the overall conclusion is that both the observed change in association and the observed shift in contribution have contributed significantly and almost equally to the decline of confessional party voting in the Netherlands between 1970 and 1998.[4]

This example was only used to clarify the method of purging on longitudinal data. However restricted to two sample points, it is quite a laborious way to produce Table II. Needless to say that purging this way is close to impossible if we use a large set of data with multiple data points and multiple causal factors. Fortunately, we do not have to recalculate the marginals this way, as there is a relationship between logistic regression analysis and the method of purging (Xie, 1989). We elaborate upon this relationship and the resulting practical applications within the field of longitudinal research in the next section.

2. The Method of Purging Using Logistic Regression Analysis

Another, more easy way to produce all relevant information from Tables I and II, is to perform a logistic regression analysis using the following equation:

$$\log \frac{p_{1(\text{conf. party voting})}}{P_{0(\text{conf. party voting})}}$$
$$= b_0 + b_1 \text{ church member } + b_2 \text{ year } + b_3 \text{ (church member} \times \text{year)} \qquad (1)$$

(coding: confessional party voting: no = 0, yes = 1; church member: non-member = 0, church member = 1; year: 1970 = 0, 1998 = 1).

Note that b_0 represents the logit among non-members in 1970, that b_1 denotes the difference in logit between non-members and church members in 1970, that b_2 is the difference in logit between 1970 and 1998 for non-members and b_3 is the additional difference in logit between 1970 and 1998 for church members. The resulting values of these parameters are presented in the second column Table III.

Next, we applied logistic regression analysis with the equation given in (1) to the data in the left part of Table I (1970) together with the data in the left part of Table II (constant association in 1998). Resulting parameters are again presented in Table III. Note that b_0 and b_1 have the same values as for Table I, since we purge on 1998 while 1970 is our base-level. The parameters b_2 and b_3 however equal zero, indicating that the association remained on the 1970 level.

Finally, a logistic regression analyses was performed on the data in the left part of Table I (1970) together with the data in right part of Table II (constant distribution in 1998) again using Equation (1). The resulting parameters are the same as for Table I, since only the marginal frequencies changed and logit parameters do not depend on marginal distributions. From the resulting parameters, we calculated the percentage that voted for a confessional party in 1998 using:

$$\hat{p}_{1(\text{conf. party voting})} = \sum_c \frac{1}{e^{-(b_0+b_1 \text{ church member}+b_2 \text{ year}+b_3 \text{ church member}\times\text{year})} + 1}$$
$$\times \frac{f_c}{f_{\text{total}}} \times 100 \qquad (2)$$

(c denotes (shaded) cell, f_c denotes the cell frequency and f_{total} is total frequency of the 1998 sample).

The outcomes are presented in the last column of Table III.

As can be seen from Table III, logistic regression provides the same percentages votes for a confessional party as in the example shown in Tables I and II. The practical implication of it all is, that we only have to perform one single logistic regression analysis on the data in Table I and let the computer calculate both the percentage confessional votes while B_2 and b_3 equal zero (= constant association) and the percentage confessional votes using the 1970 weights (f_c/f_{total}), (= constant distribution).

Table III. Percentage religious votes: (1) observed, (2) with constant association and (3) with constant distribution

Table	Parameters				Formula 2	% in 1998
	b_0	b_1	b_2	b_3		
1	−3.11	3.61	−0.09	−0.75	$1/(e^{-(-3.11+-0.09)}+1) \times 860/1495 \times 100+$	
					$1/(e^{-(-3.11+-0.09)}+1) \times 35/1495 \times 100+$	
					$1/(e^{-(-3.11+3.61+-0.09+-0.75)}+1) \times 351/1495 \times 100+$	
					$1/(e^{-(-3.11+3.61+-0.09+-0.75)}+1) \times 249/1495 \times 100 =$	19.0
Left	−3.11	3.61	0.00	0.00	$1/(e^{-(-3.11)}+1) \times 860/1495 \times 100+$	
part					$1/(e^{-(-3.11)}+1) \times 35/1495 \times 100+$	
of 2					$1/(e^{-(-3.11+3.61)}+1) \times 351/1495 \times 100+$	
					$1/(e^{-(-3.11+3.61)}+1) \times 249/1495 \times 100 =$	27.5
Right	−3.11	3.61	−0.09	−0.75	$1/(e^{-(-3.11+-0.09)}+1) \times 568/1495 \times 100+$	
part					$1/(e^{-(-3.11+-0.09)}+1) \times 23/1495 \times 100+$	
of 2					$1/(e^{-(-3.11+3.61+-0.09+-0.75)}+1) \times 375/1495 \times 100+$	
					$1/(e^{-(-3.11+3.61+-0.09+-0.75)}+1) \times 529/1495 \times 100 =$	26.6

The exact procedure to run these simulations with the SPSS package, is shown in Appendix 1.[5]

2.1. EXTENSION OF THE MODEL: MULTIPLE CAUSAL FACTORS

The model thusfar used has only two independent variables (i.e., church membership and year), whereas it may be necessary to take other variables into account as well. In our example, the association in 1970 between church membership and confessional party voting was much stronger compared to 1998. This difference mainly came about because church members turned away from confessional parties. This behavioral change may be explained by declining church attendance rates.[6] Therefore, we extend the previous model by including a variable measuring church attendance and a cross-product term modeling the effect in 1998:

$$\begin{aligned}\log \frac{p_{1(\text{rel. party voting})}}{p_{0(\text{rel. party voting})}} &= b_0 + b_1 \text{ ch. mem. } + b_2 \text{ year}\\ &\quad + b_3(\text{ch. mem. } \times \text{ year})\\ &\quad + b_4\text{ch. attend.}\\ &\quad + b_5(\text{ch. attend } \times \text{ year}) \qquad (3)\end{aligned}$$

(coding: confessional party voting: no = 0, yes = 1; church membership: non-member = 0, church member = 1; year: 0 = 1970, 1 = 1998; church attendance: 0 =$\leq$ once a month, 1 =$\geq$ once a month).

The resulting parameters are: $b_0 = -3.12$, $b_1 = 2.35$, $b_2 = -0.10$, $b_3 = -0.09$, $b_4 = 2.00$ and $b_5 = -0.45$. Compared with the previous results, b_3 is much smaller this time, indicating that church attendance (partly) explains church members' behavioral change. Next to the longitudinal change in association between church membership and religious voting (b_2 and b_3), the association between church attendance and religious voting also changed between 1970 and 1998 (b_2 and especially b_5). To assess the impact of both changes simultaneously and seperately, we carried out three simulations. In a first simulation, all longitudinal changes were purged (b_2, b_3 and b_5 set to zero). In a second simulation, only the behavioral changes of non-members and church members were purged (b_2 and b_3 set to zero). Finally, in a third simulation, the behavioral changes of church attenders and non-attenders were purged (b_2 and b_5 set to zero). The results are shown in Table IV.

The impact of a changing association is small this time, as the purged percentage differs only slightly from the observed percentage ($22.2 - 19.0 = 3.2$). Of this minor impact, only the changing association between church attendance and religious voting has a significant impact.

Likewise, one may want to know whether the decline of church membership rates or the decline of church attendance rates has had the strongest impact on

Table IV. Percentage of religious votes while simulating constant association

	% in 1998	95% confidence interval (cf. note 3)
Observed	19.0	17.0–21.0
Constant association for both		
church membership and church attendance	22.2	
Constant association for church membership	20.8	ns
Constant association for church attendance	21.5	

ns: lies within 95% confidence interval around 19%.

Table V. Percentage of religious votes while simulating constant distribution

	% in 1998	95% confidence interval (cf. note 3)
Observed	19.0	17.0—21.0
Constant distribution for both		
church membership and church attendance	33.1′	
Constant distribution for church membership	20.0	ns
Constant distribution for church attendance	23.3	

ns: lies within 95% confidence interval around 19%.

the observed loss of confessional votes. To answer this question, we used a factor decomposition method described by Liao (1989). Translated to our longitudinal problem, this method implies that for finding the impact of the shift in church membership's distribution, purged percentages must be calculated within the two categories of church attendance and multiplied by their 1998 weights of each of the church attendance categories and finally summed. To find the impact of a changing distribution of church attendance, purged percentages have to be calculated within the two categories of church membership, multiplied by their 1998 weights and summed. The decomposition method for two x-variables is expressed in the following formula:[7]

$$\hat{p}_{1(y)} = \sum_k \hat{p}_{x_k} \frac{f_{x_k}}{f_{\text{total}}} \tag{4}$$

($\hat{p}_{x_k}$ denotes the purged percentage in category k of variable x, f_{x_k} denotes the frequency in category k of variable x and f_{total} is total sample frequency).

With the use of (2), (3) and (4) the purged percentages are calculated and shown in Table V.

From Table V we derive that the constant distribution of both church membership and church attendance has a huge impact on the percentage of religious votes. If both church membership rates and church attendance would have been stable between 1970 and 1998, the purged percentage is 33.1 instead of the observed 19.0. The separate impact of each variable is modest, with the impact of falling church attendance rates being significant. The outcomes are quite logical as in the Netherlands falling church membership rates go hand in hand with falling church attendance rates (Te Grotenhuis & Scheepers, 2001). When Table V is compared with Table IV, then the conclusion must be that the dramatic loss of religious votes is primarily caused by falling church membership rates and church attendance rates. In fact, if this process would not had occurred, then the loss of confessional votes would have been 5.2% instead of the observed loss of 20.3% (cf. Table I). The exact procedure to run the simulations from Tables IV and V with the use of the SPSS package is shown in Appendix 2.

For heuristic and practical reasons the examples given were limited in data. For demonstrations of some full-fledged purging analyses (i.e., multiple factors and multiple datapoints) using logistic and linear regression analysis we like to refer to (Eisinga et al., 1997; Gomulka & Stern, 1990; Te Grotenhuis et al., 1998; Ten Have et al., 2002).

3. The Method of Purging Using Linear Regression Analysis

In the former section the dependent variable was nominal. However, the method of purging can also be applied to dependent variables which are ratio or interval scaled. As an example we choose a variable to measure respondents' opinion about abortion. From previous research it is known that church attendance and education are relative strongly related to this opinion (Scheepers et al., 2002). We therefore use the following linear regression equation with church attendance (5 categories), education (7 categories) and year (2 categories):

$$\begin{aligned}\hat{y}_{\text{abortion}} &= b_0 + b_1 \text{ ch. attendance } + b_2\text{edu. } + b_3 \text{ year}\\ &\quad + b_4(\text{ch. attendance} \times \text{year})\\ &\quad + b_5(\text{edu.} \times \text{year}) \end{aligned} \tag{5}$$

(coding: church attendance: 0 = never ... 5 = weekly; education: 0 = lowest ... 6 = highest; year: 0 = 1970, 1 = 1998).

The resulting parameters are: $b_0 = 2.60$, $b_1 = 0.27$, $b_2 = -0.10$, $b_3 = -0.15$, $b_4 = 0.07$, $b_5 = 0.034$. Note that the value of b_0 represents the mean score of respondents who do not visit religious services and have the lowest educational level in 1970, that b_1 is the effect of church attendance in 1970, that b_2 is the effect of education in 1970, that b_3 is an additional constant for 1998 while b_4 and b_5 are the additional effects of church attendance and education in 1998. Because we are only interested in the mean score on y we use:

$$\bar{y} = b_o + \sum_p b_p \bar{x}_p \tag{6}$$

($\bar{y}$ denotes mean score, b_o is constant, $\bar{x}$ is mean score on variable p).

Table VI. Mean score on abortion variable while simulating constant associations

	Mean in 1970	Mean in 1998	95% confidence interval[8]
Observed	3.02	2.86	2.80–2.92
Constant association for both church attendance and education		2.74	
Constant association for church attendance		2.87	ns
Constant association for education		2.88	ns

To hold all associations constant, b_3, b_4 and b_5 have to be set to zero in (5) and a new mean score on y is calculated. To hold all distributions constant, only the 1970 distributions are used in (5) to recalculate the new mean score on y. Following (6), this can be achieved by taking the observed means from the 1970 distributions. We can also decompose the impact of constant association and constant distribution. For instance, if one likes to know the contribution of the shift in church attendance rates, the 1998 means of year and education are substituted in (5) together with the 1970 mean of church attendance. In Table VI the outcomes of these simulations are summarized.

From Table VI we derive that between 1970 and 1998 the mean score on the abortion variable dropped from 3.02 to 2.86. This implies that, on average, Dutch people gradually opposed less to the idea that women should have the right of abortion. This decline would have been stronger if the association between church attendance and education, on the one hand, and abortion, on the other, would have been stable between 1970 and 1998. If only the association between church attendance and abortion is held constant the impact is insignificantly lower (2.87) and almost equal to a simulation in which the association between education and abortion is held constant (also non-significant). Next, we simulated constant distributions, the results can be found in Table VII.

As Table VII shows, constant distributions have quite different consequences. Unlike Table VI, the simulated mean score in 1998 has significantly risen to 3.18. This indicates that if church attendance rates would not have fallen and educational level would not have risen, the opposition against abortion would have been higher, instead of lower, compared to 1970. This impact is mainly attributable to the fall of church attendance rates (2.86–3.13) and to a much lesser degree to rising levels of education (2.86–2.91) which is inside the 95% confidence interval. The SPSS syntax to run all simulations from Tables VI and VII is shown in Appendix 3.

Table VII. Mean score on abortion variable while simulating constant distributions

	Mean in 1970	Mean in 1998	95% confidence interval (cf. note 8)
Observed	3.02	2.86	2.80–2.92
Constant distribution for both church attendance and education		3.18	
Constant distribution for church attendance		3.13	
Constant distribution for education		2.91	ns

4. To Purge or not to Purge: Discussion and Conclusions

Thusfar we assumed that the observed percentages and mean scores between 1970 and 1998 differed significantly. In other words, we assumed that there was enough statistical evidence that a social change indeed had taken place. Of course before purging is applied, one must test whether the data indicate social change. If the test fails to reject H_0 (= equal proportions or equal means) subsequent purging would have no statistical meaning.

Likewise, we assumed that the observed shift in distributions and variations of effects were significant. Of course, one should test this assumption before a purging analysis is carried out. To test whether effects changed significantly over time, it is best to include cross-product terms, modeling the interaction between the predictor variables and time (often the years in which the surveys were conducted). If these cross-product terms are not significant, purging these terms (i.e., setting them to zero) is useless. The same argument applies to shifts in distributions of predictor variables. In case there is lack of statistical evidence for this shift, one should refrain from purging them. Practically this means that one should only purge those changes and shifts that turned out to be significant.

A purging analysis is also questionable if one wants to assess the impact of a certain shift in distribution while controlling for other causal factors. This may result in a weak (controlled for) effect of the variable under investigation and hence a low purging impact even when the shift in distribution is large. We suggest to start with a simple regression analysis and assess the purging impact of a particular variable. In subsequent analyses, one may look for other variables that may explain the purging impact of that particular variable.

In this contribution we showed that the method of purging on longitudinal data can be applied with the use of logistic and linear regression analysis. As these analyses are widely used, the practical procedures we proposed may be used by a wider group of social scientists who are interested in explaining longitudinal social change. Especially in case of a large number of repeated cross-sectional surveys, we feel that the method of purging could be a powerful tool to assess to what extent distributional shifts and/or behavioural change explain longitudinal social change.

Appendix 1: SPSS Syntax for Simulation #1 and #2

```
Title  The method of purging applied to repeated cross-sectional data.
subtitle Practical applications using logistic and linear regression analysis.

* * *  Logistic regression analysis: 2 by 2 table.
data list free / chm cpv year freq .
comment:  chm=church membership, cpv=confessional party voting, year=year of
survey, freq=frequency.
begin data
0   0   0   472
0   1   0    21
1   0   0   286
1   1   0   469
0   0   1   860
0   1   1    35
1   0   1   351
1   1   1   249
end data.
weight by freq.

variable labels chm 'church member' cpv 'conf party votes'.
value labels chm 0 'none' 1 'member' / year 0 '1970' 1 '1998' / cpv 0 'no' 1
'yes'.

* * * Table 1: observed percentage confessional votes.
crosstabs tables=cpv by chm by year / cells=count, column/ sta=risk.

* * * Equation (1).
LOGISTIC REGRESSION cpv
  /METHOD=ENTER chm year chm*year.


* * * SIMULATION #1: CONSTANT ASSOCIATION.
compute pca_cpv= 1 / (exp (-1 * (-3.112 + 3.607 * chm + 0 * -0.75 * chm * year
+ 0 * -0.089 * year)) + 1) * 100.
means tables=pca_cpv by year.

* * * SIMULATION #2: CONSTANT DISTRIBUTION.
do if year eq 0.
compute year = 1.
compute pcd_cpv= 1 / (exp (-1 * (-3.112 + 3.607 * chm + 1 * -0.75 *   chm *
year  +  1 *  -0.089 * year) ) + 1) * 100.
compute year = 0.
end if.
value labels year 0 '1998'.
means tables=pcd_cpv by year.
```

Appendix 2: SPSS Syntax for Simulation #3 and #4 and Decomposition

```
Title  The method of purging applied to repeated cross-sectional data.
subtitle Practical applications using logistic and linear regression analysis.

* * *  Logistic regression analyses: 2 by 2 by 2 table.
data list free / chm cha rpv year freq .
comment:  chm=church membership, cha=chuch attendance rpv=religious party
voting,  year=year of survey, freq=frequency .

begin data
0   0   0   0   471
0   0   1   0    21
1   0   0   0   172
1   0   1   0    79
0   1   0   0     1
0   1   1   0     0
1   1   0   0   114
1   1   1   0   390
0   0   0   1   851
0   0   1   1    34
1   0   0   1   271
1   0   1   1   104
0   1   0   1     5
0   1   1   1     1
1   1   0   1    80
1   1   1   1   144
end data.
weight by freq.

variable labels chm 'church member' cha 'church attendance' rpv 'rel party
votes'.
value labels chm 0 'none' 1 'member' / cha 0 'montly or <'  1 'fortnight or >'
    / year 0  '1970' 1 '1998' / rpv  0 'no' 1 'yes'.

* * *  Observed percentage religious votes .
crosstabs tables=rpv by year / cells= count, column.

* * * Equation (3)
LOGISTIC REGRESSION rpv
  /METHOD=ENTER chm year chm*year cha cha*year.


* * * SIMULATION #3: CONSTANT ASSOCIATION.
compute pca_rpv=
1 / (exp (-1 * (-3.123 + 2.349 * chm + 0 * -0.089 * chm * year +  0 *  -0.096
* year + 2.001 * cha
+ 0 * -0.454 * cha * year ) ) + 1) * 100.
means tables=pca_rpv by year.

* * * Decompose impact of constant association.
* * * Impact of church membership.
compute pchm_rpv=
1 / (exp (-1 * (-3.123 + 2.349 * chm + 0 * -0.089 * chm * year +  0 *  -0.096
* year + 2.001 * cha
+ 1 * -0.454 * cha * year ) ) + 1) * 100.
means tables=pchm_rpv by year.

 * * Impact of church attendance.
compute pcha_rpv=
1 / (exp (-1 * (-3.123 + 2.349 * chm + 1 * -0.089 * chm * year +  0 *  -0.096
* year + 2.001 * cha
+ 0 * -0.454 * cha * year ) ) + 1) * 100.
means tables=pcha_rpv by year.


* * *  SIMULATION #4: CONSTANT DISTRIBUTION.
do if year eq 0.
compute year = 1.
compute pcd_rpv=
1 / (exp (-1 * (-3.123 + 2.349 * chm + 1 * -0.089 * chm * year +  1 *  -0.096
* year + 2.001 * cha
+ 1 * -0.454 * cha * year ) ) + 1) * 100.
compute year = 0.
end if.
value labels year 0 '1998'.
means tables=pcd_rpv by year.

* * * Decompose impact of constant distribution.
* * * Purged percentages in categories of church memberschip and church atten-
dance.

temp.
select if year eq 0.
means tables=pcd_rpv by chm.
means tables=pcd_rpv by cha.
execute.


* * *  Impact of church attendance.
do if year eq 1.
if chm eq 0 pcha=3.8700.
if chm eq 1 pcha=52.129.
end if.
des var=pcha.

* * * Impact of church membership.
do if year eq 1.
if cha eq 0 pchm=11.907.
if cha eq 1 pchm=64.195.
end if.
des var=pchm.
```

Appendix 3: SPSS Syntax for Simulation #5 and #6 and Decomposition

```
Title  The method of purging applied to repeated cross-sectional data.
subtitle Practical applications using logistic and linear regression analysis.

data list free / year freq.
begin data
0   1885
1   2097
end data.
weight by freq.

value labels  year 0  '1970' 1 '1998'.

* * * Computing observed means scores for education, church attendance and
abortion.
compute edu70=3.12.
compute chatt70=2.74.
compute abort70=3.02.
compute edu98=3.84.
compute chatt98=1.95.
compute abort98=2.86.
execute.

* * * SIMULATION #5: CONSTANT ASSOCIATION.
do if year eq 1.
compute pca= edu98 * -.098529 + chatt98 * .266171 + 0 * year * -.145825 + 0 *
chatt98 * .069640 + 0 * edu98 * .034742  + 1 * 2.600761.
end if.
execute.


* * * Decompose impact chuch attendance and education.
do if year eq 1.
compute pca_ch= edu98 * -.098529 + chatt98 * .266171 + 0 * year * -.145825 + 0
* chatt98 * .069640 + 1 * edu98 * .034742 + 1 * 2.600761.
end if.
execute


do if year eq 1.
compute pca_edu= edu98 * -.098529 + chatt98 * .266171 + 0 * year * -.145825 +
1 * chatt98 * .069640 + 0 * edu98 * .034742 + 1 * 2.600761.
end if.
execute


* * *  SIMULATION#6: CONSTANT DISTRIBUTION.
do if year eq 1.
compute pcd= edu70 * -.098529 + chatt70 * .266171 + year * -.145825 + chatt70
* .069640 + edu70 * .034742 + 1 * 2.600761.
end if.
execute.

* * * Decompose impact chuch attendance and education.
do if year eq 1.
compute pcd_ch= edu98 * -.098529 + chatt70 * .266171 + 1 * -.145825 + chatt70
* .069640 + edu98 * .034742 + 1 *  2.600761.
end if.
execute.

do if year eq 1.
compute pcd_edu= edu70 * -.098529 + chatt98 * .266171 + year * -.145825 +
chatt98 * .069640 + edu70 * .034742 + 1 * 2.600761.
end if.
execute.

split file by year.
des var=abort70 abort98 pca pca_ch pca_edu pcd pcd_ch pcd_edu.
split file off.
```

Notes

[1] In the literature on purging this simulation is called marginal CG purging and three-factor CDG purging respectively.

[2] We focused on the mainstream confessional parties, i.e., KVP, CHU, ARP and their successor CDA. Small right-wing confessional parties were not taken into account because of their stable electorate.

[3] To calculate the 95% confidence interval we used the following formula (assuming normal distribution):

$$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}} = 19 \pm 1.96\sqrt{\frac{19 \times 81}{1495}} = 19 \pm 2 = [17\%; 21\%].$$

[4] The total decline of confessional votes is 20.3% (39.3–19.0). So, there is a unexplained percentage of 4.2% (20.3–8.5–7.6). This remaining part is contributable to the joint effect of shifting distribution and changing association.

[5] All Appendices in this article can be downloaded as ASCII files from the first author's homepage (http://baserv.uci.kun.nl/~mtgroten). The files are written as syntax for SPSS for windows, version 6.0 and higher.

[6] According to social science theory, one of the explanations for this behavioral change of church members is a declining integration within a religious community (cf. Stark, 1994; Ultee et al., 1992). We take church attendance as indicator for (the lack of) integration within a religious community.

[7] The decomposition method described by Liao (1989) is not limited to two variables. One only has to calculate purged percentages within each combination of the left out variables which makes it a bit more laborious.

[8] To calculate the 95% confidence interval we used the following formula (assuming normal distribution):

$$\bar{x} \pm z_{\alpha/2}\frac{s}{\sqrt{n}} = 2.86 \pm 1.96\frac{1.3}{\sqrt{2097}} = 2.86 \pm 0.056 = [2.80; 2.92].$$

References

Clogg, C. C. (1978). Adjustment of rates using multiplicative models. *Demography* 15: 523–539.

Clogg, C. C. & Eliasen, S. R. (1988). A flexible procedure for adjusting rates and proportions, including statistical methods for group comparisons. *American Sociological Review* 53: 267–283.

Clogg, C. C., Shockey, J. W. & Eliasen, S. R. (1990). A general statistical framework for adjustment of rates. *Sociological Methods and Research* 19: 156–195.

Eisinga, R. & Felling, A. (1990). Church membership in the Netherlands, 1960–1987. *Journal for the Scientific Study of Religion* 29: 108–112.

Eisinga, R., Felling, A. & Franses, P. H. (1997). De afbrokkeling van het electoraat van (de voorlopers van) het CDA. *Sociologische Gids* 44: 77–99.

Gomulka, Joanna & Stern, Nicholas (1990). The employment of married women in the United Kingdom, 1970–83. *Economica* 57: 171–199.

Grotenhuis, M. te, Eisinga, R. & Scheepers, P. (1998). Welke gevolgen heft ontkerkelijking? Een verkenning op het terrein van cultuur, demografie, economie, politiek en welzijn in Nederland tussen 1970 en 1995. *Tijdschrift voor Sociologie* 19: 5–32.

Grotenhuis, M. te & Scheepers, P. (2001). Churches in Dutch. Causes of religious disaffiliation in the Netherlands, 1937–1995. *Journal for the Scientific Study of Religion* 40: 591–606.

Have, M. ten, Grotenhuis, M. te, Scheepers, P., Meertens, V. & Vollebergh, W. Trends in the use of community mental health and social work services in the Netherlands, 1979–1995, adjusted for behavioural changes and varying effects (submitted).

Liao, T. F. (1989). A flexible approach for the decomposition of rate differences. *Demography* 26: 717–726.

Nieuwbeerta, P. (1995). *The Democratic Class Struggle in Twenty Countries 1945–1990*. Amsterdam: Thesis Publishers

Scheepers, P., Grotenhuis, M. Te & Van Der Slik, F. (2002). Education, religiosity and moral attitudes: Explaining cross-national effect differences, *Sociology of Religion* 63: 157–176.

Stark, R. (1994). *Sociology*. Belmont: Wadsworth.

Ultee, W., Flap, H. & Arts, W. (1992). *Sociologie. Vragen, uitspraken en bevindingen*. Groningen: Wolters-Noordhoff.

Xie, Y. (1989). An alternative purging method: Controlling the composition-dependent interaction in an analysis of rates. *Demography* 26: 711–716.